Automatica 49 (2013) 3520–3530

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Efficient Bayesian spatial prediction with mobile sensor networks using Gaussian Markov random fields☆

Yunfei Xu [a], Jongeun Choi [a,b,1], Sarat Dass [c,d], Tapabrata Maiti [d]

[a] Department of Mechanical Engineering, Michigan State University, East Lansing, MI 48824, USA
[b] Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI 48824, USA
[c] Fundamental and Applied Sciences Department, Universiti Teknologi PETRONAS, Seri Iskandar, 31750 Tronoh, Perak, Malaysia
[d] Department of Statistics & Probability, Michigan State University, East Lansing, MI 48824, USA

ARTICLE INFO

Article history:
Received 30 January 2012
Received in revised form
19 February 2013
Accepted 20 August 2013
Available online 15 October 2013

Keywords:
Gaussian Markov random fields
Mobile sensor networks

ABSTRACT

In this paper, we consider the problem of predicting a large scale spatial field using successive noisy measurements obtained by mobile sensing agents. The physical spatial field of interest is discretized and modeled by a Gaussian Markov random field (GMRF) with uncertain hyperparameters. From a Bayesian perspective, we design a sequential prediction algorithm to exactly compute the predictive inference of the random field. The main advantages of the proposed algorithm are: (1) the computational efficiency due to the sparse structure of the precision matrix, and (2) the scalability as the number of measurements increases. Thus, the prediction algorithm correctly takes into account the uncertainty in hyperparameters in a Bayesian way and is also scalable to be usable for mobile sensor networks with limited resources. We also present a distributed version of the prediction algorithm for a special case. An adaptive sampling strategy is presented for mobile sensing agents to find the most informative locations in taking future measurements in order to minimize the prediction error and the uncertainty in hyperparameters simultaneously. The effectiveness of the proposed algorithms is illustrated by numerical experiments.

1. Introduction

In recent years, there has been an increasing exploitation of navigation of mobile sensor networks and robotic sensors interacting with uncertain environments (Choi, Oh, & Horowitz, 2009; Le Ny & Pappas, 2013; Leonard et al., 2007; Lynch, Schwartz, Yang, & Freeman, 2008; Stanković & Stipanović, 2010; Xu, Choi, & Oh, 2011; Zhang, Siranosian, & Krstić, 2007). A necessity in such scenarios is to design algorithms to process collected observations from environments (*e.g.*, distributed estimators) for robots such that either the local information about the environment can be used for local control actions or the global information can be estimated asymptotically. The approach to designing such algorithms takes two different paths depending on whether it uses an environmental model in space and time or not. Without environmental models, extremum seeking control has been proven to be very effective for finding the source of a signal (chemical, electromagnetic, etc.) (Stanković & Stipanović, 2010; Zhang et al., 2007). The drawback of extremum seeking control is that it limits its task to finding the maximum (or minimum) point of the environmental field. A unifying framework of distributed stochastic gradient algorithms that can deal with coverage control, spatial partitioning, and dynamic vehicle routing problems in the absence of a priori knowledge of the event location distribution, has been presented in Le Ny and Pappas (2013). However, to tackle a variety of useful tasks such as exploration, estimation, prediction and maximum seeking of a scalar field, *etc.*, it is essential for robots to have a spatial (and temporal) field model (Choi et al., 2009; Cortés, 2009; Graham & Cortés, 2009, 2012; Krause, Singh, & Guestrin, 2008; Leonard et al., 2007; Lynch et al., 2008; Varagnolo, Pillonetto, & Schenato, 2012; Xu & Choi, 2012a,b; Xu, Choi, Dass, & Maiti, 2012b; Xu et al., 2011). Although control algorithms for mobile robots have been developed based on computationally demanding, physics-based field models (*e.g.*, atmospheric dispersion modeling), recently phenomenological and statistical modeling techniques such as kriging, Gaussian processes regression, and kernel regression have gained much attention for resource-constrained mobile robots. Among phenomenological spatial models, adaptive control of multiple robotic sensors based on a parametric approach needs

☆ This work has been supported by the National Science Foundation through CAREER Award CMMI-0846547. This support is gratefully acknowledged. The material in this paper was partially presented at the 2012 American Control Conference (ACC12), June 27–29, 2012, Montreal, Canada. This paper was recommended for publication in revised form by Associate Editor Brett Ninness under the direction of Editor Torsten Söderström.

E-mail addresses: xuyunfei@egr.msu.edu (Y. Xu), jchoi@egr.msu.edu (J. Choi), sdass@msu.edu (S. Dass), maiti@msu.edu (T. Maiti).

[1] Tel.: +1 517 432 3164; fax: +1 517 353 1750.

0005-1098/$ – see front matter © 2013 Published by Elsevier Ltd
http://dx.doi.org/10.1016/j.automatica.2013.09.008

a persistent excitation (PE) condition for convergence of parameters (Choi et al., 2009), while control strategies based on Bayesian spatial models do not require such conditions, (*e.g.*, by utilizing priori distributions as in Kalman filtering (Lynch et al., 2008) or Gaussian process regression (Xu et al., 2011)). Hence, control engineers become more aware of the usefulness of nonparametric Bayesian approaches such as Gaussian processes (defined by mean and covariance functions) (Cressie, 1986; Rasmussen & Williams, 2006) to statistically model physical phenomena for the navigation of mobile sensor networks *e.g.*, Cortés (2009); Graham and Cortés (2009, 2012); Krause et al. (2008); Leonard et al. (2007); Xu and Choi (2012a); Xu et al. (2012b, 2011). Other more data-driven approaches have also developed (without a statistical structure used in Gaussian processes) such as using kernel regression (Xu & Choi, 2012b) and in reproducing kernel Hilbert spaces (Varagnolo et al., 2012). However, such an approach without a statistical structure in a random field (as in Xu and Choi (2012b), Varagnolo et al. (2012)) requires usually more observations than the one with a statistical structure for a decent prediction quality.

The significant computational complexity in Gaussian process regression due to the growing number of observations (and hence the size of covariance matrix) has been tackled in different ways (Herbrich, Lawrence, & Seeger, 2002; Seeger, 2003; Smola & Bartlett, 2001; Tresp, 2000; Williams & Seeger, 2001). In Xu et al. (2011), the authors analyzed the conditions under which near-optimal prediction can be achieved using only truncated observations when the covariance function is known *a priori*. On the other hand, unknown hyperparameters in the covariance function can be estimated by a maximum likelihood (ML) estimator or a maximum *a posteriori* (MAP) estimator and then be used in the prediction as true hyperparameters as in an empirical Bayes approach (Xu & Choi, 2011). However, the point estimate (ML or MAP estimate) itself needs to be identified using a sufficient amount of measurements and it fails to fully incorporate the uncertainty in the estimated hyperparameters into the prediction in a fully Bayesian perspective.

The advantage of a fully Bayesian approach is that the uncertainty in the model parameters are incorporated in the prediction (Bishop, 2006). However, the solution often requires Markov Chain Monte Carlo (MCMC) methods, which greatly increases the computational complexity. In Graham and Cortés (2009), an iterative prediction algorithm without resorting to MCMC methods has been developed based on analytical closed-form solutions from results in Gaudard, Karson, Linder, and Sinha (1999), by assuming that the bandwidths in the covariance function of the spatio-temporal Gaussian random field are known *a priori*. In Xu, Choi, Dass, and Maiti (2011), the authors designed a sequential Bayesian prediction algorithm to deal with unknown bandwidths by using a compactly supported kernel and selecting a subset of collected measurements. In contrast to Xu et al. (2011), in this paper, we seek a fully Bayesian approach over a discretized surveillance region such that the Bayesian spatial prediction utilizes all collected measurements in a scalable fashion.

Recently, there have been efforts to fit a computationally efficient Gaussian Markov random field (GMRF) on a discrete lattice to a Gaussian random field on a continuum space (Cressie & Verzelen, 2008; Hartman & Hössjer, 2008; Rue & Tjelmeland, 2002). It has been demonstrated that GMRFs with small neighborhoods can approximate Gaussian fields surprisingly well (Rue & Tjelmeland, 2002). This approximated GMRF and its regression are very attractive for resource-constrained mobile sensor networks due to its computational efficiency and scalability (Le Ny & Pappas, 2010) as compared to the standard Gaussian process and its regression. Fast kriging of large data sets by using a GMRF as an approximation of a Gaussian field has been proposed in Hartman and Hössjer (2008). In Xu and Choi (2012a), the authors provided a new class of Gaussian processes that builds on a GMRF and derived formulas for predictive statistics. However, they both assume the precision matrix which is the inverse of the covariance matrix is given or estimated *a priori*.

The contributions of the paper are as follows. First, we model the physical spatial field as a GMRF with uncertain hyperparameters and formulate the estimation problem from a Bayesian point of view. Second, we design an sequential Bayesian estimation algorithm to effectively and efficiently compute the exact predictive inference of the spatial field. The proposed algorithm often takes only seconds to run even for a very large spatial field, as will be demonstrated in this paper. Moreover, the algorithm is scalable in the sense that the running time does not grow as the number of observations increases. In particular, the scalable prediction algorithm does not rely on the subset of samples to obtain scalability (as was done in Xu et al. (2011)), correctly fusing all collected measurements. Thus, in contrast to previous works (Xu & Choi, 2012a; Xu et al., 2011), our prediction algorithm correctly takes into account the uncertainty in hyperparameters in a Bayesian way and is also scalable to be usable for mobile sensor networks with limited resources. We then propose a distributed prediction algorithm for a special case such that sensing agents make a prediction by exchanging only local information with their neighbors. An adaptive sampling strategy for mobile sensor networks is proposed at the end to possibly improve the quality of prediction and to reduce the uncertainty in hyperparameter estimation simultaneously. A preliminary version of this paper has been reported in Xu, Choi, Dass, and Maiti (2012a).

The paper is organized as follows. In Section 2, we introduce a spatial field model based on a GMRF, and a mobile sensor network. In Section 3, we introduce a Bayesian inference approach to estimate the spatial field of interest. A sequential Bayesian prediction algorithm is proposed in Section 4 to deal with computational complexity. A distributed prediction is then proposed in Section 5. An adaptive sampling algorithm is proposed in Section 6 for mobile sensing agents in order to minimize the prediction error and the uncertainty in hyperparameters simultaneously. We demonstrate the effectiveness through a simulation study in Section 7.

Standard notation will be used throughout the paper. Let $\mathbb{R}$, $\mathbb{R}_{>0}$, $\mathbb{Z}_{>0}$ denote, respectively, the sets of real, positive real, and positive integer numbers. The positive definiteness of a matrix A is denoted by $A \succ 0$. Let E, and Corr denote, respectively, the operators of expectation, and correlation. A random vector x, which has a multivariate normal distribution of mean vector μ and covariance matrix Σ, is denoted by $x \sim \mathcal{N}(\mu, \Sigma)$. For a set $\mathcal{A}$, the absolute complement of a subset $\mathcal{B} \subset \mathcal{A}$ is denoted by $-\mathcal{B}$. Let $\circ$ denote the element-wise product. Let $O(1)$ denote the running time of an algorithm, which is a constant. Other notation will be explained in due course.

2. Preliminaries

The objective of this paper is to design a sequential Bayesian prediction algorithm and an adaptive sampling algorithm for mobile sensing agents to consistently and accurately predict a spatial field of interest. In what follows, we specify the models for the spatial field and the mobile sensor network.

2.1. Spatial field model

Let $\mathcal{F}_* \subset \mathbb{R}^d$ denote the spatial field of interest. We discretize the field into n_* square areas, whose centers are denoted by $\mathcal{S}_* := \{s_1, \ldots, s_{n_*}\}$. Let $z_* = (z_1, \ldots, z_{n_*})^T \in \mathbb{R}^{n_*}$ be the collection of values of the field (e.g., the temperature) on the spatial sites in $\mathcal{S}_*$. Due to the irregular shape a spatial field may have, we extend the field to $\mathcal{F}$ such that $n \geq n_*$ sites denoted by $\mathcal{S} := \{s_1, \ldots, s_n\}$

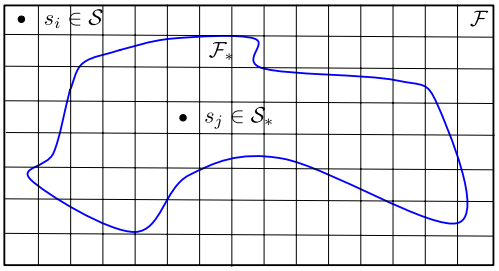


Fig. 1. Example of a spatial field. Much finer grids will be used for real applications.

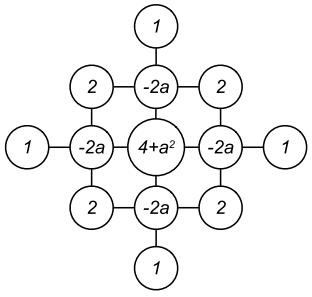


Fig. 2. Elements of the precision matrix Q related to a single location.

are on a regular grid as can be seen in Fig. 1. Notice that we have $\mathcal{S}_* \subseteq \mathcal{S}$. The latent variable $z_i = z(s_i) \in \mathbb{R}$ is modeled by

$$z_i = \mu(s_i) + \eta_i, \quad \forall 1 \le i \le n, \tag{1}$$

where $s_i \in \mathcal{S} \subset \mathbb{R}^d$ is the i-th site location. The mean function $\mu : \mathbb{R}^d \to \mathbb{R}$ is defined as

$$\mu(s_i) = f(s_i)^T \beta,$$

where $f(s_i) = (f_1(s_i), \ldots, f_p(s_i))^T \in \mathbb{R}^p$ is a known regression function, and $\beta = (\beta_1, \ldots, \beta_p)^T \in \mathbb{R}^p$ is an unknown vector of regression coefficients. We define $\eta = (\eta_1, \ldots, \eta_n)^T \in \mathbb{R}^n$ as a zero-mean Gaussian Markov random field (GMRF) (Rue & Held, 2005) denoted by

$$\eta \sim \mathcal{N}\left(0, Q_{\eta|\theta}^{-1}\right),$$

where the inverse covariance matrix (or precision matrix) $Q_{\eta|\theta} \in \mathbb{R}^{n\times n}$ is a function of a hyperparameter vector $\theta \in \mathbb{R}^m$.

There exist many different choices of the GMRF (*i.e.*, the precision matrix $Q_{\eta|\theta}$) (Rue & Held, 2005). For instance, Fig. 2 displays one choice of the elements of the precision matrix related to a single location. The associated full conditionals are shown in (2) given in Box I (with obvious notation as shown in Rue and Held (2005)). The hyperparameter vector is defined as $\theta = (\kappa, \alpha)^T \in \mathbb{R}^2_{>0}$, where $\alpha = a-4$. The resulting GMRF accurately represents a Gaussian random field with the Matérn covariance function (Lindgren, Rue, & Lindström, 2011)

$$C(r) = \sigma_f^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}r}{\ell}\right)^{\nu} K_\nu \left(\frac{\sqrt{2\nu}r}{\ell}\right),$$

where $K_\nu(\cdot)$ is a modified Bessel function (Rasmussen & Williams, 2006), with order $\nu = 1$, a bandwidth $\ell = 1/\sqrt{\alpha}$, and vertical scale $\sigma_f^2 = 1/(4\pi\alpha\kappa)$. The hyperparameter $\alpha > 0$ guarantees the positive definiteness of the precision matrix $Q_{\eta|\theta}$. In the case where $\alpha = 0$, the resulting GMRF is a second-order polynomial intrinsic GMRF (Rue & Held, 2005; Rue, Martino, & Chopin, 2009). Notice that the precision matrix is sparse which contains only a small number of non-zero elements. This property will be exploited for fast computation in the following sections.

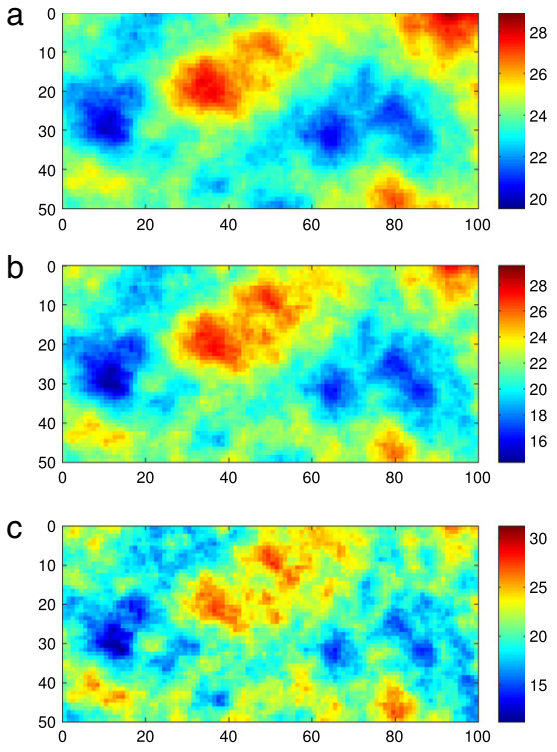


Fig. 3. Numerically generated spatial fields defined in (1) with $\mu(s_i) = \beta = 20$, and $Q_{\eta|\theta}$ constructed using (2) (see Box I) with hyperparameters being (a) $\theta = (4, 0.0025)^T$, (b) $\theta = (1, 0.01)^T$, and (c) $\theta = (0.25, 0.04)^T$.

Example 1. Consider a spatial field of interest $\mathcal{F}_* \in (0, 100) \times (0, 50)$. We first divide the spatial field into a 100×50 regular grid with equal areas 1, which makes $n_* = 5000$. We then extend the field such that 120×70 grids (*i.e.*, $n = 8400$) are constructed on the extended field $\mathcal{F} = (-10, 110) \times (-10, 60)$. The precision matrix $Q_{\eta|\theta}$ introduced above is chosen with the regular lattices wrapped on a torus (Rue & Held, 2005). (The grid is made to be wrapped on a torus such that the initial prediction error variances are the same among all grid points.) In this case, only 0.15% elements in the sparse matrix $Q_{\eta|\theta}$ are non-zero. The numerically generated fields with the mean function $\mu(s_i) = \beta = 20$, and the hyperparameter vector $\theta = (\kappa, \alpha)^T$ being different values are shown in Fig. 3.

2.2. Mobile sensor network

Consider N spatially distributed mobile sensing agents indexed by $i \in \mathcal{I} = \{1, \ldots, N\}$ sampling from n_* spatial sites in $\mathcal{S}_*$. Agents are equipped with identical sensors and sample at time $t \in \mathbb{Z}_{>0}$. At time t, agent i takes a noise corrupted measurement at its current location $q_{t,i} \in \mathcal{S}_*$, *i.e.*,

$$y_{t,i} = z(q_{t,i}) + \epsilon_{t,i}, \quad \epsilon_{t,i} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_w^2),$$

where measurement errors are assumed to be independent and identically distributed (i.i.d.). The noise level $\sigma_w^2 > 0$ is assumed to be known. For notational simplicity, we denote all agents' locations at time t by $q_t = (q_{t,1}^T, \ldots, q_{t,N}^T)^T$ and the observations made by all agents at time t by $y_t = (y_{t,1}, \ldots, y_{t,N})^T$. Furthermore, we denote the collection of agents' locations and the collective observations from time 1 to t by $q_{1:t} = (q_1^T, \ldots, q_t^T)^T$, and $y_{1:t} = (y_1^T, \ldots, y_t^T)^T$, respectively.

3. Bayesian predictive inference

In this section, we propose a Bayesian inference approach to make predictive inferences of a spatial field $z_* \in \mathbb{R}^{n_*}$.

$$\mathrm{E}(\eta_i|\eta_{-i},\theta)=\frac{1}{4+a^2}\begin{pmatrix}\circ&\circ&\circ&\circ&\circ&\circ&\circ&\circ&\circ&\circ&\circ&\circ&\bullet&\circ&\circ\\ \circ&\circ&\bullet&\circ&\circ&\circ&\bullet&\circ&\bullet&\circ&\circ&\circ&\circ&\circ&\circ\\ 2a\circ&\bullet&\circ&\bullet&\circ&-2\circ&\circ&\circ&\circ&\circ&-1\bullet&\circ&\circ&\circ&\bullet\\ \circ&\circ&\bullet&\circ&\circ&\circ&\bullet&\circ&\bullet&\circ&\circ&\circ&\circ&\circ&\circ\\ \circ&\circ&\circ&\circ&\circ&\circ&\circ&\circ&\circ&\circ&\circ&\circ&\bullet&\circ&\circ\end{pmatrix},\tag{2}$$

$$\mathrm{Var}(\eta_i|\eta_{-i},\theta)=(4+a^2)\kappa$$

Box I.

First, we assign the vector of regression coefficients $\beta\in\mathbb{R}^p$ with a Gaussian prior, namely $\beta\sim\mathcal{N}\left(0,T^{-1}\right)$, where the precision matrix $T\in\mathbb{R}^{p\times p}$ is often chosen as a diagonal matrix with small diagonal elements when no prior information is available. Hence, the distribution of latent variables z given β and the hyperparameter vector θ is Gaussian, *i.e.*,

$$z|\beta,\theta\sim\mathcal{N}\left(F\beta,Q_{\eta|\theta}^{-1}\right),$$

where $F=(f(s_1),\ldots,f(s_n))^T\in\mathbb{R}^{n\times p}$. For notational simplicity, we denote the full latent field of dimension $n+p$ by $x=(z^T,\beta^T)^T$. Then, for a given hyperparameter vector θ, the distribution $\pi(x|\theta)$ is Gaussian, obtained by

$$\begin{aligned}\pi(x|\theta)&=\pi(z|\beta,\theta)\pi(\beta)\\&\propto\exp\left(-\frac{1}{2}(z-F\beta)^TQ_{\eta|\theta}(z-F\beta)-\frac{1}{2}\beta^TT\beta\right)\\&=\exp\left(-\frac{1}{2}x^TQ_{x|\theta}x\right),\end{aligned}$$

where the precision matrix $Q_{x|\theta}\in\mathbb{R}^{(n+p)\times(n+p)}$ is defined by

$$Q_{x|\theta}=\begin{pmatrix}Q_{\eta|\theta}&-Q_{\eta|\theta}F\\-F^TQ_{\eta|\theta}&F^TQ_{\eta|\theta}F+T\end{pmatrix}.$$

By the matrix inversion lemma, the covariance matrix $\Sigma_{x|\theta}\in\mathbb{R}^{(n+p)\times(n+p)}$ can be obtained by

$$\Sigma_{x|\theta}=Q_{x|\theta}^{-1}=\begin{pmatrix}Q_{\eta|\theta}^{-1}+FT^{-1}F^T&FT^{-1}\\(FT^{-1})^T&T^{-1}\end{pmatrix}.$$

At time $t\in\mathbb{Z}_{>0}$, we have a collection of observational data $y_{1:t}\in\mathbb{R}^{Nt}$ obtained by the mobile sensing agents over time. Let $A_{1:t}=(A_1,\ldots,A_t)\in\mathbb{R}^{(n+p)\times Nt}$, where $A_\tau\in\mathbb{R}^{(n+p)\times N}$ is defined by

$$(A_\tau)_{ij}=\begin{cases}1,&\text{if } i\le n \text{ and } s_i=q_{\tau,j},\\0,&\text{otherwise.}\end{cases}$$

Then the covariance matrix of $y_{1:t}$ can be obtained by

$$R_{1:t}=A_{1:t}^T\Sigma_{x|\theta}A_{1:t}+P_{1:t},$$

where $P_{1:t}=\sigma_w^2I\in\mathbb{R}^{Nt\times Nt}$. By Gaussian process regression (Rasmussen & Williams, 2006), the full conditional distribution of x is also Gaussian, *i.e.*,

$$x|\theta,y_{1:t}\sim\mathcal{N}(\mu_{x|\theta,y_{1:t}},\Sigma_{x|\theta,y_{1:t}}),\tag{3}$$

where

$$\begin{aligned}\Sigma_{x|\theta,y_{1:t}}&=\Sigma_{x|\theta}-\Sigma_{x|\theta}A_{1:t}R_{1:t}^{-1}A_{1:t}^T\Sigma_{x|\theta},\\\mu_{x|\theta,y_{1:t}}&=\Sigma_{x|\theta}A_{1:t}R_{1:t}^{-1}y_{1:t}.\end{aligned}\tag{4}$$

The posterior distribution of the hyperparameter vector θ can be obtained via

$$\pi(\theta|y_{1:t})\propto\pi(y_{1:t}|\theta)\pi(\theta),$$

where the log likelihood function is defined by

$$\log\pi(y_{1:t}|\theta)=-\frac{1}{2}y_{1:t}^TR_{1:t}^{-1}y_{1:t}-\frac{1}{2}\log\det R_{1:t}-\frac{Nt}{2}\log2\pi.\tag{5}$$

If a discrete prior on the hyperparameter vector θ is chosen with a support $\Theta=\{\theta_1,\ldots,\theta_L\}$, the posterior predictive distribution $\pi(x|y_{1:t})$ can be obtained by

$$\pi(x|y_{1:t})=\sum_\ell\pi(x|\theta_\ell,y_{1:t})\pi(\theta_\ell|y_{1:t}).\tag{6}$$

The predictive mean and variance then follow as

$$\begin{aligned}\mu_{x_i|y_{1:t}}&=\sum_\ell\mu_{x_i|\theta_\ell,y_{1:t}}\pi(\theta_\ell|y_{1:t}),\\\sigma_{x_i|y_{1:t}}^2&=\sum_\ell\sigma_{x_i|\theta_\ell,y_{1:t}}^2\pi(\theta_\ell|y_{1:t})\\&\quad+\sum_\ell(\mu_{x_i|\theta_\ell,y_{1:t}}-\mu_{x_i|y_{1:t}})^2\pi(\theta_\ell|y_{1:t}),\end{aligned}\tag{7}$$

where $\mu_{x_i|\theta_\ell,y_{1:t}}$ is the i-th element in $\mu_{x|\theta_\ell,y_{1:t}}$, and $\sigma_{x_i|\theta_\ell,y_{1:t}}^2$ is the i-th diagonal element in $\Sigma_{x|\theta_\ell,y_{1:t}}$.

Remark 2. The discrete prior $\pi(\theta)$ greatly reduces the computational complexity in that it enables summation in (6) instead of numerical integration which has to be performed for a continuous prior distribution. However, the computation of the full conditional distribution $\pi(x|\theta,y_{1:t})$ in (4) and the likelihood $\pi(y_{1:t}|\theta)$ (5) still requires the inversion of the covariance matrix $R_{1:t}$, whose size grows as time t increases. Thus, the running time grows as new observations are collected and it will soon become intractable. This problem will be addressed in the following sections thanks to the sparsity structure of the precision matrix brought by the GMRF model.

4. Sequential Bayesian inference

In this section, we exploit the sparsity of the precision matrix, and derive a sequential Bayesian prediction algorithm which can be performed in constant time as the number of observations increases.

4.1. Sequential update on full conditional distribution

First, we derive the update rule for the full conditional distribution in (3). From here on, we will use $Q_{t|\theta}:=Q_{x|\theta,y_{1:t}}$ and $\mu_{t|\theta}:=\mu_{x|\theta,y_{1:t}}$, for notational simplicity. Moreover, we define $u_{t,i}\in\mathbb{R}^{n+p}$ as

$$(u_{t,i})_j=\begin{cases}1,&\text{if } j\le n \text{ and } s_j=q_{t,i},\\0,&\text{otherwise.}\end{cases}$$

We have the following propositions.

Proposition 3. *The full conditional distribution in (3) can be obtained by*

$$x|\theta,y_{1:t}\sim\mathcal{N}(Q_{t|\theta}^{-1}b_t,Q_{t|\theta}^{-1}),\tag{8}$$

where

$$\begin{aligned}Q_{t|\theta}&=Q_{t-1|\theta}+\frac{1}{\sigma_w^2}\sum_{i=1}^N u_{t,i}u_{t,i}^T,\\b_t&=b_{t-1}+\frac{1}{\sigma_w^2}\sum_{i=1}^N u_{t,i}y_{t,i},\end{aligned}\tag{9}$$

with initial conditions

$$Q_{0|\theta} := Q_{x|\theta,y_{1:0}} = Q_{x|\theta}, \quad \text{and} \quad b_0 = 0.$$

Proof. The result followed by the Woodbury identity and simple algebra.

Proposition 4. *The full conditional mean and variance can be updated by*

$$\mu_{t|\theta} = Q_{t|\theta}^{-1} b_t,$$
$$\operatorname{diag}(Q_{t|\theta}^{-1}) = \operatorname{diag}(Q_{t-1|\theta}^{-1}) - \sum_{i=1}^{N} \frac{h_{t,i|\theta} \circ h_{t,i|\theta}}{\sigma_w^2 + u_{t,i}^T h_{t,i|\theta}}, \tag{10}$$

where

$$h_{t,i|\theta} = B_{t,i|\theta}^{-1} u_{t,i},$$
$$B_{t,i|\theta} = Q_{t-1|\theta} + \frac{1}{\sigma_w^2} \sum_{j=1}^{i} u_{t,j} u_{t,j}^T.$$

Proof. The result followed by the Sherman–Morrison formula and simple algebra.

We then have the following lemma.

Lemma 5. *For a given $\theta \in \Theta$, the full conditional mean and variance, i.e., $\mu_{t|\theta}$ and $\operatorname{diag}(Q_{t|\theta}^{-1})$ in (10), can be updated in $O(1)$ given $Q_{t-1|\theta}$ and b_{t-1}.*

Proof. The update of $Q_{t|\theta}$ and b_t can be obviously computed in constant time. Then $\mu_{t|\theta}$ in (10) can be obtained by solving a linear equation $Q_{t|\theta}\mu_{t|\theta} = b_t$. Due to the sparse structure of $Q_{t|\theta}$, this operation can be done in a very short time. Moreover, notice that $Q_{t|\theta}$ and $Q_{t-1|\theta}$ have the same sparsity structure and hence the computational complexity remains fixed. Similarly, $h_{t,i|\theta}$ can be obtained in $O(1)$ and hence $\operatorname{diag}(Q_{t|\theta}^{-1})$ in (10) can be obtained in $O(1)$.

4.2. Sequential update on likelihood

Next, we derive the update rule for the log likelihood function. We have the following proposition.

Proposition 6. *The log likelihood function $\log \pi(y_{1:t}|\theta)$ in (5) can be obtained by*

$$\log \pi(y_{1:t}|\theta) = c_t + g_{t,\theta} + \frac{1}{2} b_t^T \mu_{t|\theta} - \frac{Nt}{2} \log(2\pi\sigma_w^2), \tag{11}$$

where

$$c_t = c_{t-1} - \frac{1}{2\sigma_w^2} \sum_{i=1}^{N} y_{t,i}^2, \quad c_0 = 0,$$
$$g_{t|\theta} = g_{t-1|\theta} - \frac{1}{2} \sum_{i=1}^{N} \log\left(1 + \frac{1}{\sigma_w^2} u_{t,i}^T h_{t,i|\theta}\right), \quad g_{0|\theta} = 0,$$

with $h_{t,i|\theta}$ defined in (10).

Proof. See Appendix.

The computation of the likelihood function is scalable as follows.

Lemma 7. *For a given $\theta \in \Theta$, the log likelihood function, i.e., $\log \pi(y_{1:t}|\theta)$ can be updated in $O(1)$.*

Proof. The result follows directly from Proposition 6.

4.3. Sequential update on predictive distribution

Combining the results in Lemmas 5, 7, and (6), (7), we summarize our results in the following theorem.

Theorem 8. *The predictive distribution in (6) (or the predictive mean and variance in (7)) can be obtained in $O(1)$ as time t increases.*

The proposed sequential Bayesian prediction algorithm is summarized in Table 1.

5. Distributed prediction algorithm

In this section, we first briefly review the distributed computation algorithms. We then propose a distributed approach to effectively implement the sequential Bayesian predictive algorithm for a special case.

5.1. Distributed computation

In this subsection, we briefly review two useful methods to convert the centralized sequential prediction algorithm in Table 1 into a distributed algorithm for a mobile sensor network.

5.1.1. Jacobi over-relaxation method

The Jacobi over-relaxation (JOR) (Bertsekas & Tsitsiklis, 1999) method provides an iterative solution of a linear system $Ax = b$, where $A \in \mathbb{R}^{N\times N}$ is a nonsingular matrix and $x, b \in \mathbb{R}^N$. If agent i knows the $\text{row}_i(A) \in \mathbb{R}^N$ and b_i, and $a_{ij} = (A)_{ij} = 0$ if agent i and agent j are not neighbors, then the recursion is given by

$$x_i^{(k+1)} = (1-h)x_i^{(k)} + \frac{h}{a_{ii}} \left(b_i - \sum_{j\in\mathcal{N}_i} a_{ij} x_j^{(k)}\right). \tag{12}$$

This JOR algorithm converges to the solution of $Ax = b$ from any initial condition if $h < 2/N$ (Cortés, 2009). At the end of the algorithm, agent i knows the i-th element of $x = A^{-1}b$.

In the case where the size of the matrix A is much larger than the number of agents N, and each agent knows multiple rows of A, a similar algorithm can be derived.

5.1.2. Discrete-time average consensus

The discrete-time average consensus (DAC) provides a way to compute the arithmetic mean of elements in the vector $c \in \mathbb{R}^N$. Assume the graph is connected. If agent i knows the i-th element of c, the network can compute the arithmetic mean via the following recursion (Olfati-Saber, Fax, & Murray, 2007)

$$x_i^{(k+1)} = x_i^{(k)} + \epsilon \sum_{j\in\mathcal{N}_i} a_{ij}(x_j^{(k)} - x_i^{(k)}), \tag{13}$$

with initial condition $x(0) = c$, where $a_{ij} = 1$ if j and i are neighbors and 0 otherwise, $0 < \epsilon < 1/\Delta$, and $\Delta = \max_i(\sum_{j\neq i} a_{ij})$ is the maximum degree of the network. After the algorithm converges, all nodes in the network know the average of c, *i.e.*, $\sum_{i=1}^n c_i/N$.

In the case where the size of c is much larger than the number of agents N, and each agent knows multiple elements of c, a summation within each agent can be performed beforehand.

5.2. Distributed implementation

To design a distributed version of the proposed algorithm, we make the following assumptions for a robotic sensor network with limited resources:

A1. The mean values of the field are known to be $\mu(s_i) = 0$, *i.e.*, $p = 0$ and $Q_{x|\theta} = Q_{\eta|\theta} \in \mathbb{R}^{n\times n}$.

A2. The extended spatial field containing n spatial sites is partitioned into N subregions. Each agent is in charge of one subregion as shown in Fig. 4. Notice that each agent can only move within a part of the subregion that belongs to the real spatial field $\mathcal{F}_*$.

Table 1
Sequential Bayesian predictive algorithm.

Input: (1) prior distribution of $\theta \in \Theta$, *i.e.*, $\pi(\theta)$
(2) spatial sites $\mathcal{S}_* := \{s_1, \cdots, s_{n_*}\}$
(3) extended sites $\mathcal{S} := \{s_1, \cdots, s_n\}$
(4) regression function $f(\cdot)$

Output: (1) predictive mean $\mu_{x_i|y_{1:t}}$
(2) predictive variance $\sigma^2_{x_i|y_{1:t}}$

Initialization:
1: initialize $b = 0 \in \mathbb{R}^{n+p}$, $u = 0 \in \mathbb{R}^{n+p}$, $c = 0 \in \mathbb{R}$
2: **for** $\theta \in \Theta$ **do**
3: initialize $Q_\theta \in \mathbb{R}^{(n+p)\times(n+p)}$, $g_\theta = 0 \in \mathbb{R}$
4: compute $\mathrm{diag}(Q_\theta^{-1}) \in \mathbb{R}^{n+p}$
5: **end for**

At time $t \in \mathbb{Z}_{>0}$, do:
1: **for** $1 \le i \le N$ **do**
2: obtain new observations $y_{t,i}$ collected at current locations $q_{t,i}$
3: find the index k corresponding to $q_{t,i}$, and set $(u)_k = 1$
4: update $(b)_k = (b)_k + y_{t,i}/\sigma_w^2$
5: update $c = c - y_{t,i}^2/(2\sigma_w^2)$
6: **for** $\theta \in \Theta$ **do**
7: compute $h_\theta = Q_\theta^{-1}u$
8: update $\mathrm{diag}(Q_\theta^{-1}) = \mathrm{diag}(Q_\theta^{-1}) - \frac{h_\theta \circ h_\theta}{\sigma_w^2 + u^T h_\theta}$
9: update Q_θ via $(Q_\theta)_{kk} = (Q_\theta)_{kk} + 1/\sigma_w^2$
10: update $g_\theta = g_\theta - \frac{1}{2}\log(1 + \frac{1}{\sigma_w^2}u^T h_\theta)$
11: **end for**
12: **end for**
13: **for** $\theta \in \Theta$ **do**
14: compute $\mu_\theta = Q_\theta^{-1}b$
15: compute the likelihood via
$$\log \pi(y_{1:t}|\theta) = c + g_\theta + \tfrac{1}{2}b^T\mu_\theta$$
16: **end for**
17: compute the posterior distribution via
$$\pi(\theta|y_{1:t}) \propto \pi(y_{1:t}|\theta)\pi(\theta)$$
18: compute the predictive mean via
$$\mu_{x_i|y_{1:t}} = \textstyle\sum_\ell (\mu_{\theta_\ell})_i \pi(\theta_\ell|y_{1:t})$$
19: compute the predictive variance via
$$\sigma^2_{x_i|y_{1:t}} = \textstyle\sum_\ell \left((\mathrm{diag}(Q_{\theta_\ell}))_i + ((\mu_{\theta_\ell})_i - \mu_{x_i|y_{1:t}})^2\right)\pi(\theta_\ell|y_{1:t})$$

A3. Agent i knows certain rows of Q and elements of b, denoted by $Q^{[i]}$ and $b^{[i]}$ respectively, the indices of which correspond to the spatial sites within its subregion. If agent i and agent j are not neighbors, then the sub-matrix $Q^{[i][j]}$ is zero. N agents collectively have the knowledge of all rows of Q and all n elements of b.

A4. Each agent can only communicate with its neighbors.

The distributed prediction algorithm can be obtained similarly as the centralized algorithm (Table 1). In the distributed algorithm, the linear equations in Steps 7 and 14 can be solved in a distributed fashion by using the Jacobi over-relaxation method (Section 5.1.1) with assumption *A4*. In Step 15, the third term in the expression of the likelihood function can be easily evaluated using the discrete-time average consensus algorithm described in Section 5.1.2. At the end of the algorithm (Steps 18 and 19), each agent obtains the mean and variance on the sites within its own subregion.

Remark 9. By assumptions *A2* and *A3*, each agent is in charge of the computation for the spatial sites within its own subregion.

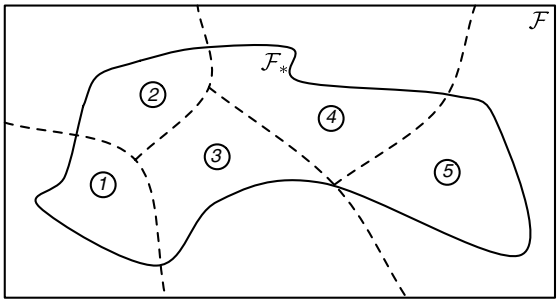


Fig. 4. Partition of the extended field Q.

Therefore, the partition of the field determines the computational power required for each agent. The partitioning could be created according to the agents' computational capability, for all agents to finish each step in more or less same time. Such an intelligent partitioning could be done by a load balancing algorithm in a distributed manner, for example, one as shown in Durham, Carli, Frasca, and Bullo (2009). In this case, if an agent is equipped with a more powerful computer than others, then a larger subregion would be assigned to this agent.

6. Adaptive sampling

In the previous section, we have designed a sequential Bayesian prediction algorithm and its distributed version for estimating the scalar field at time t. In this section, we propose an adaptive sampling strategy for finding most informative sampling locations at time $t+1$ for mobile sensing agents in order to improve the quality of prediction and reduce the uncertainty in hyper parameters simultaneously.

In our previous work (Xu et al., 2011), we have proposed using the conditional entropy $H(z_*|\theta = \hat{\theta}_t, y_{1:t+1})$ as an optimality criterion, where $\hat{\theta}_t = \arg\max_\theta \pi(\theta|y_{1:t})$ is the maximum *a posteriori* (MAP) estimate based on the cumulative observations up to current time t. Although this approach greatly simplifies the computation, it does not count for the uncertainty in estimating the hyperparameter vector θ.

In this paper, we propose to use the conditional entropy $H(z_*, \theta|\tilde{y}, y_{1:t})$ which represents the uncertainty remaining in both random vectors z_* and θ by knowing future measurements in the random vector $\tilde{y}$. Notice that the measurements $y_{1:t}$ have been observed and treated as constants. It can be obtained by

$$\begin{aligned} H(z_*, \theta|\tilde{y}, y_{1:t}) &= H(z_*|\theta, \tilde{y}, y_{1:t}) + H(\theta|\tilde{y}, y_{1:t}) \\ &= H(z_*|\theta, \tilde{y}, y_{1:t}) + H(\tilde{y}|\theta, y_{1:t}) \\ &\quad + H(\theta|y_{1:t}) - H(\tilde{y}|y_{1:t}). \end{aligned}$$

Notice that we have the following Gaussian distributions (the means will not be exploited and hence not shown here):

$$z_*|\theta, \tilde{y}, y_{1:t} \sim \mathcal{N}(\cdot, \Sigma_{z_*|\theta,\tilde{y},y_{1:t}}),$$
$$\tilde{y}|\theta, y_{1:t} \sim \mathcal{N}(\cdot, \Sigma_{\tilde{y}|\theta,y_{1:t}}),$$
$$\tilde{y}|y_{1:t} \overset{\text{approx}}{\sim} \mathcal{N}(\cdot, \Sigma_{\tilde{y}|y_{1:t}}),$$

in which the last one is approximated by a Gaussian distribution with mean and variance computed as in (7). Moreover, the entropy $H(\theta|y_{1:t}) = c$ is a constant since $y_{1:t}$ is known. Since the entropy for a multivariate Gaussian distribution has a closed-from expression (Cover & Thomas, 2006), we have

$$\begin{aligned} H(z_*, \theta|\tilde{y}, y_{1:t}) &= \sum_\ell \frac{1}{2}\log\left((2\pi e)^{n_*}\det(\Sigma_{z_*|\theta_\ell,\tilde{y},y_{1:t}})\right)\pi(\theta_\ell|y_{1:t}) \\ &\quad + \sum_\ell \frac{1}{2}\log\left((2\pi e)^{N}\det(\Sigma_{\tilde{y}|\theta_\ell,y_{1:t}})\right)\pi(\theta_\ell|y_{1:t}) \\ &\quad - \frac{1}{2}\log\left((2\pi e)^{N}\det(\Sigma_{\tilde{y}|y_{1:t}})\right) + c. \end{aligned}$$

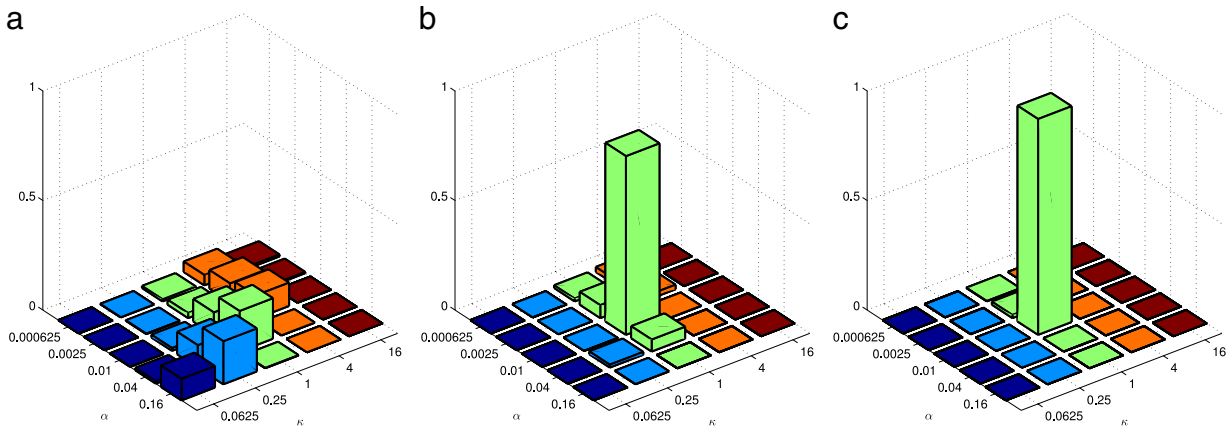


Fig. 5. Posterior distributions of θ, *i.e.*, $\pi(\theta|y_{1:t})$, at (a) $t = 1$, (b) $t = 5$, and (c) $t = 20$ using $H(z_*, \theta|\tilde{y}, y_{1:t})$ as the optimality criterion.

It can also be shown that

$$\log\det(\Sigma_{z_*|\theta_\ell,\tilde{y},y_{1:t}}) = \log\det(M^{-1}_{(\mathcal{S}_*)}) = \log\det(M_{(-\mathcal{S}_*)}) - \log\det(M),$$

where $M = Q_{t+1|\theta_\ell}$ and $M_{(\mathcal{S}_*)}$ denotes the submatrix of M formed by the first 1 to n_* rows and columns (recall that $\mathcal{S}_* = \{s_1, \ldots, s_{n_*}\}$). Notice that the term $\log\det(Q_{t+1|\theta_\ell})_{(-\mathcal{S}_*)}$ is a constant since agents only sample at $\mathcal{S}_*$. Hence, the optimal sampling locations at time $t+1$ can be determined by solving the following optimization problem

$$\begin{aligned} q_{t+1} &= \arg\min_{\{\tilde{q}_i \in \mathcal{R}_{t,i}\}} H(z_*, \theta|\tilde{y}, y_{1:t}) \\ &= \arg\min_{\{\tilde{q}_i \in \mathcal{R}_{t,i}\}} \sum_\ell -\log\det(Q_{t+1|\theta_\ell})\pi(\theta_\ell|y_{1:t}) \\ &\quad + \sum_\ell \log\det(\Sigma_{\tilde{y}|\theta_\ell,y_{1:t}})\pi(\theta_\ell|y_{1:t}) - \log\det(\Sigma_{\tilde{y}|y_{1:t}}), \end{aligned}$$

where $\mathcal{R}_{t,i} = \{s \mid \|s - q_{t,i}\| \le r, s \in \mathcal{S}_*\}$ (in which $r \in \mathbb{R}_{>0}$ is the maximum distance an agent can move between time instances) is the reachable set at time t. This combinatorial optimization problem can be solved using a greedy algorithm, *i.e.*, finding the suboptimal sampling locations for agents in sequence.

7. Simulation

In this section, we demonstrate the effectiveness of the proposed sequential Bayesian inference algorithm and the adaptive sampling strategy through a numerical experiment.

Consider a spatial field introduced in Example 1. The mean function is a constant $\beta = 20$. We choose the precision matrix $Q_{x|\theta}$ with hyperparameters $\alpha = 0.01$ equivalent to a bandwidth $\ell = 1/\sqrt{\alpha} = 10$, and $\kappa = 1$ equivalent to a vertical scale $\sigma_f^2 = 1/(4\pi\alpha\kappa) \approx 8$. The numerically generated field is shown in Fig. 6(a). The precision matrix T of β is chosen to be 10^{-4}. The measurement noise level $\sigma_w = 0.2$ is assumed to be known. The support of the prior distribution is chosen to be $\alpha \in \{0.000625, 0.0025, 0.01, 0.04, 0.16\}$ and $\kappa \in \{0.0625, 0.25, 1, 4, 16\}$. Notice that the support for α is chosen such that the bandwidth is contained within 2.5 and 40, which is very likely for a field $\mathcal{F}$ of size 100×50. And the selected κ together with α cover a huge selection of σ_f^2. The discrete uniform prior distribution is chosen. $N = 5$ mobile sensing agents take measurements at time $t \in \mathbb{Z}_{>0}$, starting from locations shown in Fig. 6(b) (as white dots). The maximum distance each agent can travel between time instances is chosen to be $r = 5$.

Fig. 6 shows the predicted fields and the prediction error variances at times $t = 1, 5, 20$. It can be seen that agents try to cover the field of interest as time evolves. The predicted field (the predictive mean) gets closer to the true field (see Fig. 6(a)) and the prediction error variances become smaller as more observations are collected. Fig. 5 shows the posterior distribution of the hyperparameters in θ. Clearly, as more measurements are obtained, this posterior distribution becomes peaked at the true value $(1, 0.01)$. Fig. 7(a) shows the predicted distribution of the estimated mean β as time evolves. In Fig. 7(b), we can see that the RMS error computed via

$$\mathrm{rms}(t) = \sqrt{\frac{1}{n_*}\sum_{i=1}^{n_*}(\mu_{z_i|y_{1:t}} - z_i)^2},$$

decreases as time increases, which shows the effectiveness of the proposed scheme.

Remark 10. For comparison, we also ran the same simulation under the same conditions except using $H(z_*|\hat{\theta}_t, \tilde{y}, y_{1:t})$ as the optimality criterion instead. The estimated β and the RMS error are shown in Fig. 9. The estimated θ at time $t = 5$ is shown in Fig. 8. Compared to Fig. 5(b), we can see that the results using $H(z_*, \theta|\tilde{y}, y_{1:t})$ as the optimality criterion are slightly better in the sense that θ converges to the true value faster. The RMS errors between the two criteria are not much different for our simulation setup as illustrated by Figs. 7(b) and 9(b). These observations from the two simulation results are expected since both the criteria minimize the uncertainty in z_* while the proposed criterion $H(z_*, \theta|\tilde{y}, y_{1:t})$ minimizes the uncertainty in θ in addition. Using $H(z_*, \theta|\tilde{y}, y_{1:t})$ could potentially be more useful when the uncertainty level in θ is large.

To show the robustness of the proposed algorithm with respect to the selection of the support (the discrete prior distribution), we implemented the algorithm using exactly same support for the hyperparameters. Notice that this time the field is generated with $\alpha = 0.02$ and $\kappa = 0.5$ which are not on the discrete support. The root mean square error is shown in Fig. 10. As can be seen, even if the true hyperparameter vector is not included in the discrete prior, the prediction is comparable.

Next, we applied our algorithm on a general case in which the true field is generated according to

$$z(j) = 20 + 10\exp\left(\frac{\|s(j) - [30, 20]^T\|}{20}\right) + 8\exp\left(\frac{\|s(j) - [80, 40]^T\|}{15}\right).$$

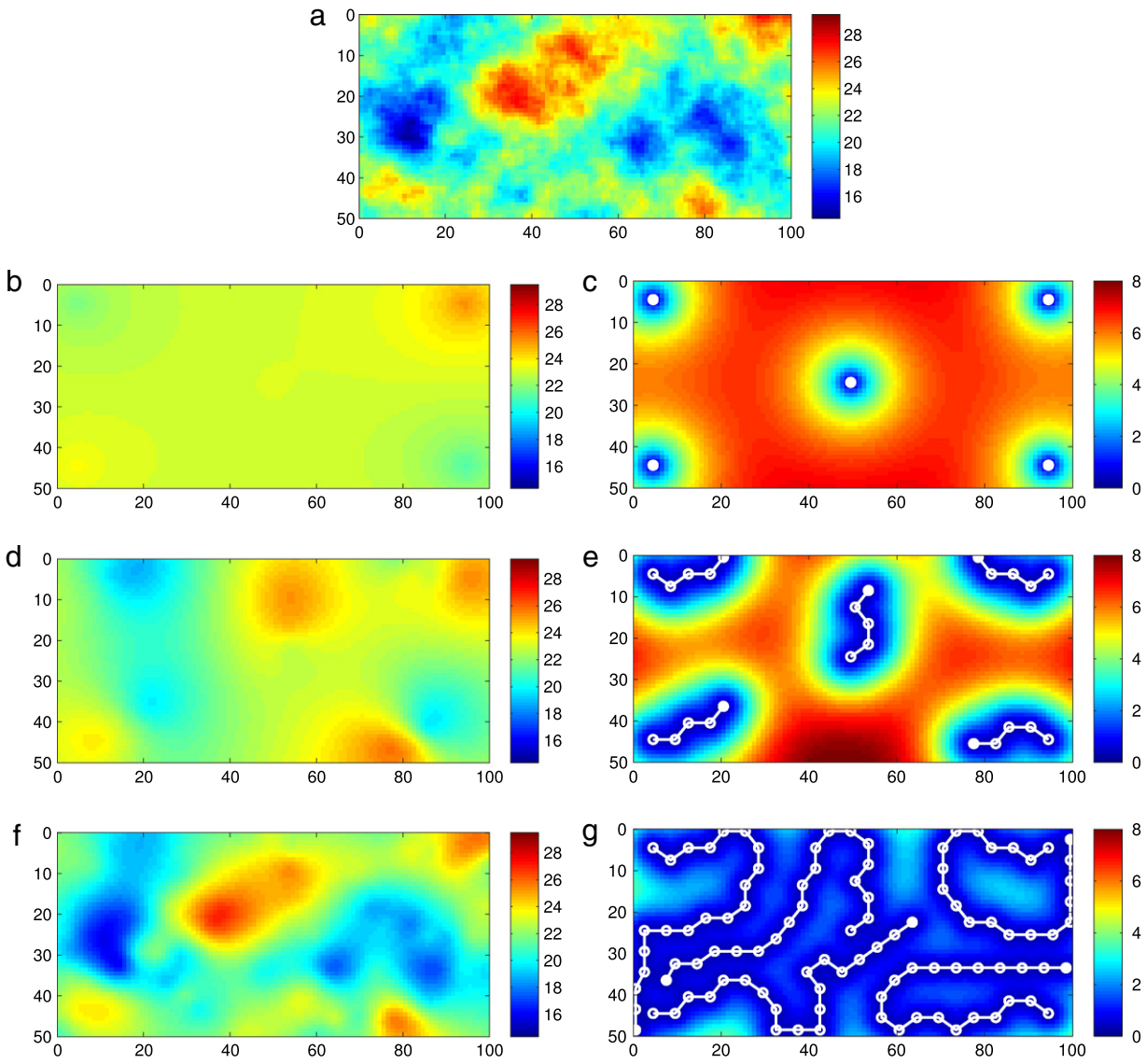


Fig. 6. The true field is shown in (a). Predicted fields at (b) $t = 1$, (d) $t = 5$, and (f) $t = 20$. Prediction error variances at (c) $t = 1$, (e) $t = 5$, and (g) $t = 10$. This result is obtained by using the optimality criterion $H(z_*, \theta|\tilde{y}, y_{1:t})$. The trajectories of agents are shown as white circles with the current locations shown as white dots.

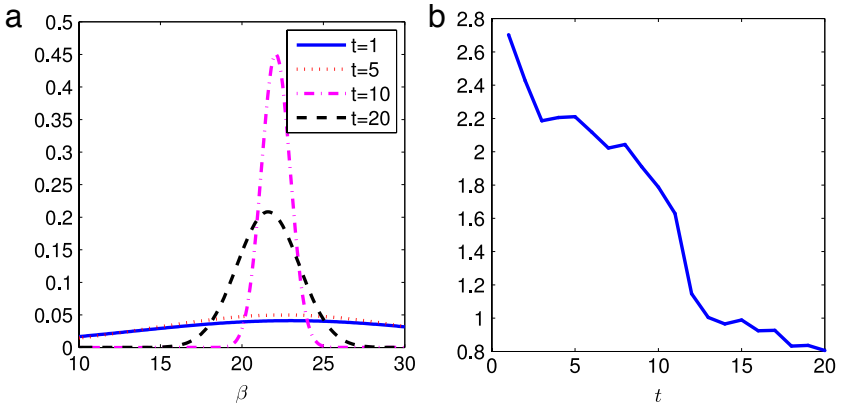


Fig. 7. (a) Estimated β, and (b) root mean square error, using $H(z_*, \theta|\tilde{y}, y_{1:t})$ as the optimality criterion.

As we can see in Fig. 11, the prediction is reasonably good with the same discrete prior distribution.

The most important contribution is that the computation time at each time step does not grow as the number of measurements

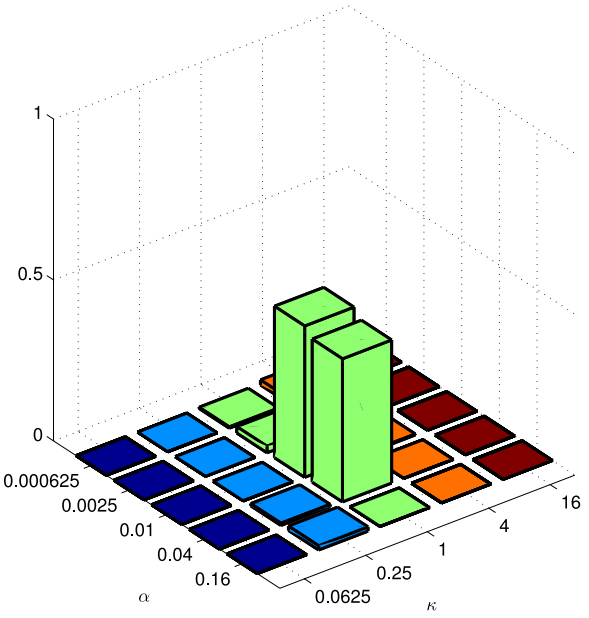


Fig. 8. Posterior distributions of θ, *i.e.*, $\pi(\theta|y_{1:t})$, at $t = 5$ using $H(z_*|\hat{\theta}_t, \tilde{y}, y_{1:t})$ as the optimality criterion.

increases. For this illustrative simulation example, this fixed running time using Matlab, R2009b (MathWorks) on a Mac (2.4 GHz Intel Core 2 Duo Processor) is about 30 s which is fast enough for real-world implementation.

8. Conclusion

We have discussed the problem of predicting a large scale spatial field using successive noisy measurements obtained by a multi-agent system. We modeled the spatial field of interest using a GMRF and designed a sequential prediction algorithm for computing the exact predictive inference from a Bayesian point of view. The proposed algorithm is computationally efficient and scalable as the number of measurements increases. A distributed prediction algorithm was also developed for a special case such that the prediction can be computed without a central station. We designed an adaptive sampling algorithm for agents to find the sub-optimal locations in order to minimize the prediction error and reduce the uncertainty in hyperparameters simultaneously. The illustrative simulation results suggested that our computationally efficient algorithms can be used for robotic sensors under realistic situations, e.g., large surveillance regions. Future work will consider the approximate Bayesian inference with a continuous prior on the hyperparameter vector.

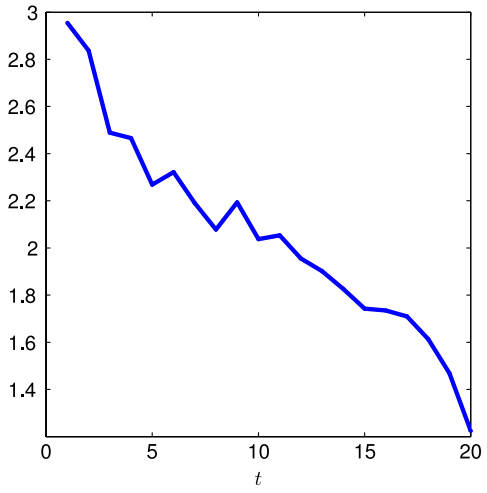


Fig. 10. The root mean square error for the robustness test, using $H(z_*|\hat{\theta}_t, \tilde{y}, y_{1:t})$ as the optimality criterion.

Acknowledgments

The authors thank the associate editor and the anonymous reviewers for their valuable comments and suggestions.

Appendix. Proof of Proposition 6

Proof. The inverse of the covariance matrix $R_{1:t}$ can be obtained by

$$\begin{aligned} R_{1:t}^{-1} &= (A_{1:t}^T Q_{0|\theta}^{-1} A_{1:t} + P_{1:t})^{-1} \\ &= P_{1:t}^{-1} - P_{1:t}^{-1} A_{1:t}^T (Q_{0|\theta} + A_{1:t} P_{1:t}^{-1} A_{1:t}^T)^{-1} A_{1:t} P_{1:t}^{-1} \\ &= P_{1:t}^{-1} - P_{1:t}^{-1} A_{1:t}^T Q_{t|\theta}^{-1} A_{1:t} P_{1:t}^{-1}. \end{aligned}$$

Similarly, the log determinant of the covariance matrix $\Sigma_{1:t}$ can be obtained by

$$\begin{aligned} \log\det R_{1:t} &= \log\det(A_{1:t}^T Q_{0|\theta}^{-1} A_{1:t} + P_{1:t}) \\ &= \log\det\left(I + \frac{1}{\sigma_w^2} A_{1:t}^T Q_{0|\theta}^{-1} A_{1:t}\right) + Nt\log\sigma_w^2 \\ &= \log\det\left(Q_{0|\theta} + \frac{1}{\sigma_w^2}\sum_{\tau=1}^{t}\sum_{i=1}^{N} u_{\tau,i} u_{\tau,i}^T\right) \\ &\quad - \log\det(Q_{0|\theta}) + Nt\log\sigma_w^2 \\ &= \sum_{\tau=1}^{t} \log(1 + u_\tau^T Q_{\tau-1|\theta}^{-1} u_\tau) + Nt\log\sigma_w^2. \end{aligned}$$

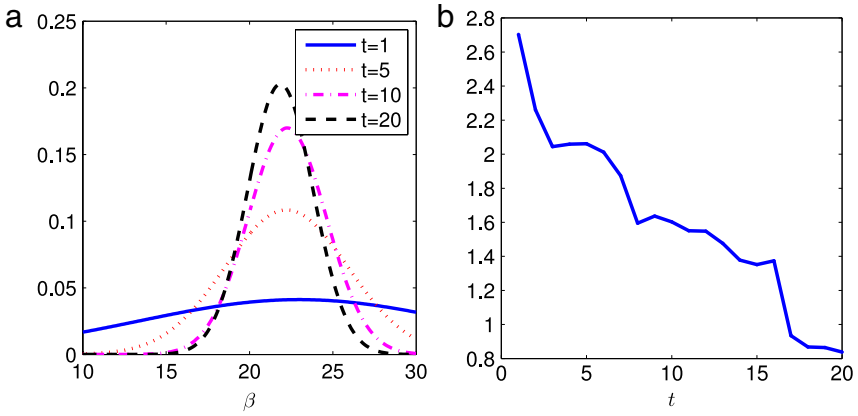


Fig. 9. (a) Estimated β, and (b) root mean square error, using $H(z_*|\hat{\theta}_t, \tilde{y}, y_{1:t})$ as the optimality criterion.

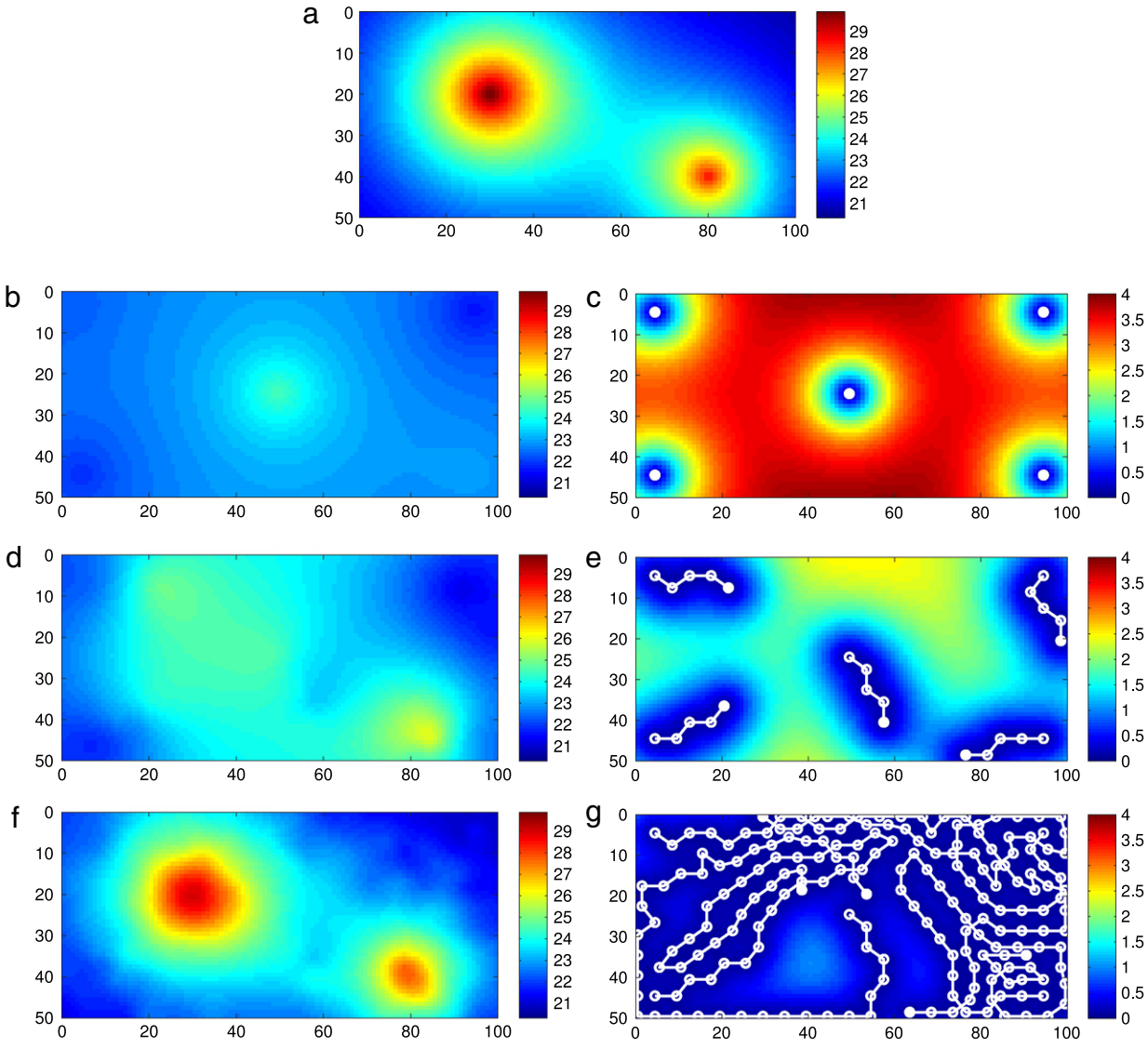


Fig. 11. The simulation results for a general field. True field is shown in (a). Predicted fields at (b) $t = 1$, (d) $t = 5$, and (f) $t = 40$. Prediction error variances at (c) $t = 1$, (e) $t = 5$, and (g) $t = 40$. This result is obtained by using the optimality criterion $H(z_*, \theta|\bar{y}, y_{1:t})$. The trajectories of agents are shown as white circles with the current locations shown as white dots.

Hence, we have

$$\begin{aligned}\log \pi(y_{1:t}|\theta) &= -\frac{1}{2}y_{1:t}^T R_{1:t}^{-1} y_{1:t} - \frac{1}{2}\log\det R_{1:t} - \frac{Nt}{2}\log 2\pi \\ &= -\frac{1}{2}y_{1:t}^T P_{1:t}^{-1} y_{1:t} + \frac{1}{2}b_t^T \mu_{t|\theta} \\ &\quad - \frac{1}{2}\sum_{\tau=1}^{t}\sum_{i=1}^{N}\log(1 + u_{\tau,i}^T B_{\tau,i|\theta}^{-1} u_{\tau,i}) \\ &\quad - \frac{Nt}{2}\log(2\pi\sigma_w^2),\end{aligned}$$

which completes the proof.

References

Bertsekas, D. P., & Tsitsiklis, J. N. (1999). *Parallel and distributed computation: numerical methods*. Englewood Cliffs: Prentice Hall.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer.

Choi, J., Oh, S., & Horowitz, R. (2009). Distributed learning and cooperative control for multi-agent systems. *Automatica*, *45*, 2802–2814.

Cortés, J. (2009). Distributed kriged Kalman filter for spatial estimation. *IEEE Transactions on Automatic Control*, *54*(12), 2816–2827.

Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.

Cressie, N. (1986). Kriging nonstationary data. *Journal of the American Statistical Association*, *81*(395), 625–634.

Cressie, N., & Verzelen, N. (2008). Conditional-mean least-squares fitting of Gaussian Markov random fields to Gaussian fields. *Computational Statistics & Data Analysis*, *52*(5), 2794–2807.

Durham, J.W., Carli, R., Frasca, P., & Bullo, F. (2009). Discrete partitioning and coverage control with gossip communication. In *ASME dynamic systems and control conference* (pp. 225–232).

Gaudard, M., Karson, M., Linder, E., & Sinha, D. (1999). Bayesian spatial prediction. *Environmental and Ecological Statistics*, *6*(2), 147–171.

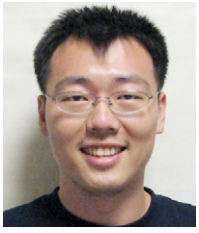
Graham, R., & Cortés, J. (2009). Cooperative adaptive sampling of random fields with partially known covariance. *International Journal of Robust and Nonlinear Control*, *1*, 1–2.

Graham, R., & Cortés, J. (2012). Adaptive information collection by robotic sensor networks for spatial estimation. *IEEE Transactions on Automatic Control*, *57*(6), 1404–1419.

Hartman, L., & Hössjer, O. (2008). Fast kriging of large data sets with Gaussian Markov random fields. *Computational Statistics & Data Analysis*, *52*(5), 2331–2349.

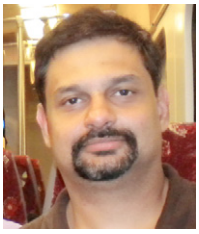
Herbrich, Ralf, Lawrence, Neil D, & Seeger, Matthias (2002). Fast sparse Gaussian process methods: the informative vector machine. *Advances in Neural Information Processing Systems*, 609–616.

Krause, A., Singh, A., & Guestrin, C. (2008). Near-optimal sensor placements in Gaussian processes: theory, efficient algorithms and empirical studies. *The Journal of Machine Learning Research*, *9*, 235–284.
Le Ny, J., & Pappas, G.J. (2010). On trajectory optimization for active sensing in Gaussian process models. In *Proceedings of the 48th IEEE conference on decision and control* (pp. 6286–6292).
Le Ny, J., & Pappas, G. J. (2013). Adaptive deployment of mobile robotic networks. *IEEE Transactions on Automatic Control*, *58*(3), 654–666.
Leonard, N. E., Paley, D. A., Lekien, F., Sepulchre, R., Fratantoni, D. M., & Davis, R. E. (2007). Collective motion, sensor networks, and ocean sampling. *Proceedings of the IEEE*, *95*(1).
Lindgren, F., Rue, H., & Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B*, *73*(4), 423–498.
Lynch, K. M., Schwartz, I. B., Yang, P., & Freeman, R. A. (2008). Decentralized environmental modeling by mobile sensor networks. *IEEE Transactions on Robotics*, *24*(3), 710–724.
Olfati-Saber, R., Fax, J. A., & Murray, R. M. (2007). Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE*, *95*(1), 215–233.
Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. Cambridge, Massachusetts, London, England: The MIT Press.
Rue, H., & Held, L. (2005). *Gaussian Markov random fields: theory and applications*. Chapman & Hall.
Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *71*(2), 319–392.
Rue, H., & Tjelmeland, H. (2002). Fitting Gaussian Markov random fields to Gaussian fields. *Scandinavian Journal of Statistics*, *29*(1), 31–49.
Seeger, Matthias (2003). *Bayesian Gaussian process models: PAC-Bayesian generalisation error bounds and sparse approximations* (Ph.D. thesis). School of Informatics, University of Edinburgh.
Smola, A. J., & Bartlett, P. (2001). Sparse greedy Gaussian process regression. *Advances in Neural Information Processing Systems*, *13*.
Stanković, M. S., & Stipanović, D. M. (2010). Extremum seeking under stochastic noise and applications to mobile sensors. *Automatica*, *46*(8), 1243–1251.
Tresp, V. (2000). A Bayesian committee machine. *Neural Computation*, *12*(11), 2719–2741.
Varagnolo, D., Pillonetto, G., & Schenato, L. (2012). Distributed parametric and nonparametric regression with on-line performance bounds computation. *Automatica*, *48*(10), 2468–2481.
Williams, C. K. I., & Seeger, M. (2001). Using the Nyström method to speed up kernel machines. *Advances in Neural Information Processing Systems*, *13*.
Xu, Y., & Choi, J. (2011). Adaptive sampling for learning Gaussian processes using mobile sensor networks. *Sensors*, *11*(3), 3051–3066.
Xu, Y., & Choi, J. (2012a). Spatial prediction with mobile sensor networks using Gaussian processes with built-in Gaussian Markov random fields. *Automatica*, *48*(8), 1735–1740.
Xu, Y., & Choi, J. (2012b). Stochastic adaptive sampling for mobile sensor networks using kernel regression. *International Journal of Control, Automation and Systems*, *10*(4), 778–786.
Xu, Y., Choi, J., Dass, S., & Maiti, T. (2011) Bayesian prediction and adaptive sampling algorithms for mobile sensor networks. In *Proceedings of the 2011 American control conference (ACC)* (pp. 4095–4200). San Francisco, CA.
Xu, Y., Choi, J., Dass, S., & Maiti, T. (2012a). Efficient Bayesian spatial prediction with mobile sensor networks using Gaussian Markov random fields. In *American control conference (ACC), 2012* (pp. 2171–2176).
Xu, Y., Choi, J., Dass, S., & Maiti, T. (2012b). Sequential Bayesian prediction and adaptive sampling algorithms for mobile sensor networks. *IEEE Transactions on Automatic Control*, *57*(8), 2078–2084.
Xu, Y., Choi, J., & Oh, S. (2011). Mobile sensor network navigation using Gaussian processes with truncated observations. *IEEE Transactions on Robotics*, *27*(6), 1118–1131.
Zhang, C., Siranosian, A., & Krstić, M. (2007). Extremum seeking for moderately unstable systems and for autonomous vehicle target tracking without position measurements. *Automatica*, *43*(10), 1832–1839.



Yunfei Xu received his Ph.D. in Mechanical Engineering from Michigan State University in 2011. He also received his M.S. and B.S. degrees in Automotive Engineering from Tsinghua University, Beijing, China, in 2004 and 2007, respectively. Currently, he is a Research Fellow with the Department of Mechanical Engineering at Michigan State University. His current research interests include environmental adaptive sampling algorithms, Gaussian processes, and statistical learning algorithms with applications to robotics and mobile sensor networks. His papers were finalists for the Best Student Paper Award at the Dynamic System and Control Conference (DSCC) 2011 and 2012.

Jongeun Choi received his Ph.D. and M.S. degrees in Mechanical Engineering from the University of California at Berkeley in 2006 and 2002 respectively. He also received a B.S. degree in Mechanical Design and Production Engineering from Yonsei University at Seoul, Republic of Korea in 1998. He is currently an Associate Professor with the Departments of Mechanical Engineering and Electrical and Computer Engineering at the Michigan State University. His current research interests include systems and control, system identification, and Bayesian methods, with applications to mobile robotic sensors, environmental adaptive sampling, engine control, neuromusculoskeletal systems, and biomedical problems. He is an Associate Editor for the Journal of Dynamic Systems, Measurement and Control. His papers were finalists for the Best Student Paper Award at the 24th American Control Conference (ACC) 2005 and the Dynamic System and Control Conference (DSCC) 2011 and 2012. He was a recipient of an NSF CAREER Award in 2009. Dr. Choi is a member of ASME.



Sarat Dass received his Ph.D. and M.S. degrees in Statistics from Purdue University at West Lafayette, Indiana, US, in 1995 and 1998, respectively. He also received a B.Stat. (Hons) degree in Statistics from the Indian Statistical Institute in 1993. His current research interests include statistical inference for dynamical systems, statistical pattern recognition and image processing, and Bayesian methods with applications to various fields of engineering. He is an Associate Editor for Sankhya B, the Journal of the Indian Statistical Institute. He received the Outstanding Statistical Application award in 2012 from the American Statistical Association (ASA) for his paper entitled "Assessing Fingerprint Individuality: A Case Study in Spatially Dependent Marked Processes". The paper also received the Wilcoxon award from Technometrics. Dr. Dass is a member of ASA and ISBA.

Tapabrata Maiti is a world class statistician, a fellow of the American Statistical Association and the Institute of Mathematical Statistics. He has published research articles in top tier statistics journals such as the journal of the American Statistical Association, Annals of Statistics, the Journal of the Royal Statistical Society, Series B, Biometrika, Biometrics, etc. He has also published research articles in engineering, economics, genetics, medicine and social sciences. His research has been supported by the National Science Foundation and National Institutes of Health. He presented his work in numerous national and international meetings and in academic departments. Prof. Maiti served in editorial board of several statistics journals including journal of the American Statistical Association and journal of Agricultural, Environmental and Biological Statistics. He has also served on several professional committees. Currently, he is a professor and the graduate director in the department of statistics and probability, Michigan State University. Prior to MSU, he was a tenured faculty member in the department of statistics, Iowa State University. Professor Maiti has supervised several Ph.D. students and regularly teaches statistics and non-stat major graduate students.